STATE OF CALIFORNIA
DEPARTMENT OF MOTOR VEHICLES

# EVALUATION OF THE SPANISH CLASS C DRIVER LICENSE WRITTEN KNOWLEDGE TESTS

**April 2008**

**Research and Development Branch**
**Licensing Operations Division**

**Author: Sukhvir S. Brar**



**RSS-08-226**

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188



| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | April 2008 | Final Report |

4. TITLE AND SUBTITLE
Evaluation of the Spanish Class C Driver License Written Knowledge Tests

5. FUNDING NUMBERS

6. AUTHOR(S)
Sukhvir S. Brar

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
California Department of Motor Vehicles
Research and Development Section
P.O. Box 932382
Sacramento, CA 94232-3820

8. PERFORMING ORGANIZATION REPORT NUMBER
CAL-DMV-RSS-08-226

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report presents the results of an evaluation of the Spanish language written knowledge tests completed by applicants for an original or renewal Class C driver license. The report presents the test fail rate, mean number of errors, and internal-consistency reliability for each test form, as well as the pass rate, item choice selection rates, and item-total correlation for each question on each form. Items that need to be reviewed for possible rewording or replacement are identified. The results are based on 4,539 completed test forms collected from all California Department of Motor Vehicle field offices from July 9 through 13, 2007.

14. SUBJECT TERMS
written test, fail rates, reliability, item analysis

15. NUMBER OF PAGES
39

16. PRICE CODE

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | None |

# PREFACE

This report is issued as an internal monograph of the Department of Motor Vehicles' Research and Development Branch rather than an official report of the State of California. The findings and opinions may not represent the views and policies of the State of California.

ACKNOWLEDGMENTS

This project was conducted under the general direction of David DeYoung, Research Chief, and the supervision of Robert Hagge, Research Manager II. Special thanks to Christine Rose, Program Support Staff Manager V, Herbert Brown, Program Support Staff Manager III, Martha Lewis, Driver License Manager IV, Jose Garcia, Associate Governmental Program Analyst, and the Spanish Translation Team for their contributions to the successful completion of this study. Staff in the Driver License Issuance Unit, under the supervision of Gerald Armstrong, Manager I, key entered the data from the completed written tests. Debbie McKenzie, Associate Governmental Program Analyst, helped type and proofread the report.

# EXECUTIVE SUMMARY

## Introduction

- This report presents the results of an evaluation of the Spanish DL 5 (Rev. 6/07) Class C license written knowledge examinations. The study assessed the fail rate, mean number of errors, and internal-consistency reliability for each test form, as well as the pass rate, percentage of applicants selecting each answer choice, and item-total correlation for each item on each test form.

- The results are based on 4,539 completed test forms that were collected from field offices from July 9 through 13, 2007.

## Results

- The overall fail rates for applicants on the first test attempt are 73.4% for original applicants and 71.6% for renewal applicants. The fail rates for both groups tend to remain steady or increase over the first three attempts.

- The first-attempt fail rates for individual test forms were fairly consistent for original applicants, ranging from 68.7% to 75.8%. Except for Form 4, which has a fail rate of 54.4%, the first-attempt fail rates are consistently high for renewal applicants on the first test attempt, ranging from 74.7% to 79.3%.

- For first-attempt originals, all test reliability coefficients are above the .70 standard of acceptability, with values ranging from .77 to .84. For first-attempt renewals, the coefficient values range from .48 to .72, with three forms falling below the .70 standard.

- Several of the items on each test form are potentially deficient as indicated by their having low item-total correlations, item pass rates that are too high or too low, or distracter selection rates that are too high or too low.

- Examiners exercised their discretion in rephrasing missed test questions and changed some applicants' original answer choices accordingly. This resulted in the

examiner fail rate being lower than the computer graded fail rate by 2.0 percentage points for originals and 9.3 percentage points for renewals on the first test attempt.

- Some field offices administered the back of the DL 5 test sheet to renewal applicants, which is not consistent with the department's policy of using only the front of the sheet for renewals.

## Recommendations

- The test questions with statistical characteristics that indicate they may be deficient should be reviewed and modified or replaced as necessary. This should include a review of language translation. Any revision of the test questions should be done without changing the equivalency of knowledge content between the English and Spanish versions of the tests.

- The order of correct answer choices within each test item should be periodically randomized to guard against rote memorization of the correct answer sequence on older revisions of the tests.

- Renewal applicants should be required to complete all 36 items on the DL 5. This will raise the reliability of the renewal tests above the .70 standard.

- The department should discontinue the policy that allows examiners to restate missed test questions. This will increase the integrity of the testing process by removing examiner subjectivity from scoring.

- Steps should be taken to ensure that field office personnel are administering only the current version of each test.

- The department should continue investigating the possible use of computer technology to automate the construction and administration of the knowledge tests.

- The following strategies should be considered for increasing applicant awareness and knowledge, thereby improving test scores:

    1. The department should prepare and distribute informational brochures and press releases that publicize the specific knowledge content areas and principles

that are most challenging to applicants, and the importance of reading the driver handbook before taking the test.

2. Procedures should be developed and implemented to identify low-literacy applicants who need to take the knowledge test orally rather than in a written form.

3. The Spanish version of the *California Drivers Handbook* should be made more widely available and distributed in both written and electronic forms.

4. A minimum waiting period should be required between test attempts to increase the likelihood that applicants will study the *California Drivers Handbook* more thoroughly before taking additional tests. The waiting period would also be an incentive to study before the first test attempt to avoid repeat trips to the field office. At a minimum, applicants should not be given multiple tests on the same day.

# TABLE OF CONTENTS



## APPENDICES

TABLE OF CONTENTS (continued)

LIST OF TABLES



LIST OF FIGURES

# INTRODUCTION

This report presents the results of an evaluation of the Spanish DL 5 (Rev. 6/07) written knowledge examinations required of applicants for an original or renewal non-commercial Class C driver license. These tests measure an individual's knowledge of traffic laws and safe-driving practices covered in the *California Driver's Handbook*. The test and item statistics from this evaluation are meant to be used to guide any further revisions of the test forms to improve their reliability and validity.

The Spanish DL 5 test is administered to Class C license applicants who request a written driver knowledge test in Spanish. These applicants may study for the test by reading the Spanish translation of the *California Driver's Handbook*. The test and handbook are also available in 34 other non-English languages. Applicants who are not literate may take an oral exam in lieu of the written test.

There are five different versions or forms of the Spanish DL 5 examination. Each form consists of 36 questions. These forms are direct translations of the first five forms of the English DL 5 test; the remaining five forms of the English DL 5 are not translated into Spanish. These tests are administered to Class C applicants who are 18 years of age or older. (Applicants under the age of 18 take a 46-item test and are not included in this study.) Original license applicants complete all 36 items, while license renewal applicants complete only the first 18 items. All test questions are multiple choice with three possible answers. Original applicants must answer 30 out of 36 items correctly to pass, while renewal applicants must answer 15 out of 18 items correctly.

The entire item pool for the English DL 5 is comprised of 342 questions that were developed by knowledge experts in the department. The items cover only subject matter in the *California Driver's Handbook* and represent 23 different knowledge content areas. Items from each content area were assigned to 10 forms of the English DL 5 in equal proportions. Since the five forms of the Spanish DL 5 are direct translations of the first five forms of the English DL 5, they also represent each content area in equal

proportions. The content areas and the number of unique items assigned to each test form are shown in Table 1.

This report presents the fail rate, mean number of errors, and internal-consistency reliability for each form of the Spanish DL 5 tests completed by each applicant group on their first test attempt. In addition, test fail rates are presented for first, second, third, and fourth or higher attempts on all forms combined for each applicant group. Finally, the pass rate, percentage of applicants choosing each answer choice, item-total correlation for each item on each test form, and the percentage of applicants who would pass at different cut-points, are presented for each applicant group on the first test attempt.

The individual questions on the Spanish DL 5 had not been analyzed in recent years and therefore it was previously unknown whether the same questions that are problematic for English-language applicants are also problematic for applicants who take the test in Spanish, or vice versa. The current evaluation analyzed the Spanish translation of the questions, so that items that are problematic can be identified and corrected.

Table 1

Knowledge Content Areas Covered by
Items on the Spanish DL 5 (Rev. 6/07) Written Tests

| Knowledge content area | Total number of items | Number of items on each test form |
|---|---|---|
| Accident responsibility | 5 | 1 |
| Driving under the influence | 5 | 1 |
| Driving on freeways | 5 | 1 |
| Lane markings | 5 | 1 |
| Lane usage | 5 | 1 |
| Mandatory questions (BAC & vehicle sale) | 2 | 2 |
| Parking on hills | 5 | 1 |
| Parking in general | 5 | 1 |
| Road hazards | 5 | 1 |
| Railroad crossings | 5 | 1 |
| Right-of-way | 10 | 2 |
| Seat belts and child restraints | 5 | 1 |
| Space cushion around vehicle | 10 | 2 |
| Speed and speed limits | 10 | 2 |
| Safe driving practices | 15 | 3 |
| Sharing the roadway with others | 5 | 1 |
| Driving special vehicles | 5 | 1 |
| Improving traffic flow | 5 | 1 |
| Traffic lights and signals | 10 | 2 |
| Turns | 10 | 2 |
| Traffic signs | 15 | 3 |
| Visual scanning | 15 | 3 |
| Driving in inclement weather | 10 | 2 |
| Total | 172 | 36 |

*Note.* There are five equivalent forms of the Spanish DL5 written test. Each form contains one or more items that relate to the interaction of vehicles and pedestrians. The items are typically drawn from either the right-of-way, safe driving practices, sharing the roadway with others, traffic lights and signals, or visual scanning content areas. The "safe driving practices" category contains items relating to vehicle equipment usage (e.g., horn, headlights, turn signals, parking lights, and emergency flashers), general safe driving rules, accident avoidance and protection, defensive driving, driving when tired, and other subject matter.

# METHODS

## Data Collection

All DMV field offices were asked to send to the department's Research and Development Branch (R&D) all Spanish non-commercial Class C driver license written knowledge tests completed from July 9 through 13, 2007. Six field offices were subsequently excluded from performing this task due to their involvement in another DMV pilot being conducted at the same time. Of the remaining participating offices, 156 submitted usable data and 8 did not. The latter offices either did not implement data collection or submitted older, unusable revisions of the tests.

Copies of the most current revision of the Spanish DL 5 were sent to the field offices for use in the evaluation. The examiners were asked to grade each test in the normal way by circling the correct answer for each missed question and writing the number of missed questions at the top of the test sheet. They were also asked not to tear off the test form number or give the test back to the customer, and to record the following information on the front of each test form: the test administration date, field office number, applicant's driver license number, test attempt number (1st, 2nd, 3rd, etc.), and whether the customer was an original ("ORIG") or renewal ("REN") applicant.

## Data Analysis

Analysis of data was conducted using the Statistical Package for the Social Sciences (SPSS). SPSS was initially used to determine the number of questions that were answered incorrectly on each test form and whether each applicant passed or failed the test. A statistical technique known as analysis of variance (ANOVA) was used to determine whether any of the differences between test form fail rates or average scores (number of errors) are statistically significant. Differences are considered to be statistically significant if the probability ($p$) of their occurrence by chance alone is less than 5 times in 100. If significant differences were found by the omnibus ANOVA,

Games and Howell multiple-comparisons tests were then used to determine what specific rates or means significantly differ from one another.

The internal-consistency reliability of each test form was computed using the Kuder-Richardson formula (KR-20). This type of reliability indicates the degree of uniformity in subject matter content among test items, and commensurately the overall precision of the test as a measurement instrument. If a test has high reliability, a person should achieve a similar test score over repeated administrations of the test (assuming that the person's true knowledge level does not change between the tests). The reliability coefficient can range in value from 0 to 1. A value of 0 indicates that no similarity exists between the test items. A value of 1, on the other hand, denotes that the items on the test are perfectly homogeneous in content. Thus, coefficients closer to 1 indicate greater reliability and are more desirable. Psychometricians generally agree that the reliability of a written knowledge test should be at least .70.

The item-total correlation coefficient measures the strength of the relationship between performance on an individual test item and performance on the overall test. The coefficient can range from -1 to 1. Items with a positive correlation coefficient are more likely to be answered correctly by applicants with high test scores, whereas items with a negative coefficient are more likely to be answered correctly by applicants with low test scores. A coefficient value close to zero indicates that answering the test question correctly or incorrectly has very little or no relationship with whether an applicant scores high or low on the overall test, which may be due to wording ambiguity or some other problem with the question.

# RESULTS

## Test Data Collection and Screening

A total of 4,591 completed test forms were received by R&D. Of these, 52 were excluded from the evaluation because they were not the most recent revision of the Spanish DL 5, were renewal tests in which the back of the test sheet was administered instead of the front, were administered outside the data collection period, lacked critical information, or were completed by an applicant who was caught cheating. It is very unlikely that the exclusion of these forms biased the results because so few were eliminated, and because the reasons they were excluded are not likely related to test performance. Data from the remaining 4,539 usable tests were key-entered into a Microsoft Access database by the department's Driver Licensing Issuance Unit.

A question was scored as incorrect if a wrong answer choice was selected, the item was left blank, or more than one answer choice was marked. The total test scores from computer grading were used to compute the test form fail rates, mean number of errors, and internal-consistency reliabilities. The fail rates are based on the current passing standards that allow no more than six errors for original applicants and three errors for renewal applicants. At least 100 first-attempt test sheets are necessary to produce reasonably precise test and item statistics for a given test form. This standard was met for all forms for each applicant group.

## Test Statistics

*Examiner Scoring Bias*

Previous evaluations of the DMV's written knowledge tests have shown that the fail rates resulting from DMV examiner grading are substantially lower than those based on computer grading. This occurs because DMV policy allows examiners to rephrase or restate missed questions on renewal law tests and give credit for those the customer ultimately answers correctly. To determine the extent of examiner scoring bias, two different fail rates were computed for test sheets that had an examiner score or test result (pass or fail) recorded (4.8% of all first attempt tests for both groups). One fail rate is based on computer grading and the other is based on the scores recorded by examiners. Because it is not known whether examiners who rephrased missed test questions consistently wrote the score on the test sheet, the extent to which these results

reflect the impact of rephrasing missed questions is not completely clear. Based on these data, Table 2 shows the resultant fail rates for originals and renewals on the first test attempt.

Compared to the computer graded fail rate (reflecting the applicants original answer choice), the fail rate based on the examiners' scores was 2.0 percentage points lower for originals and 9.3 percentage points lower for renewals. These results indicate that the examiners exercised the policy of allowing restating or rephrasing questions for renewals and that this substantially affected the test result and ultimate licensing decision in favor of these applicants. In addition, the results also show that the examiners were using this practice for originals, which this policy does not permit.

Table 2

Number of Tests (*n*) and Fail Rates for Original and Renewal Applicants on the First Test Attempt When Tests Were Graded by Computer Versus Examiner

| Applicant group | *n* | Computer fail rate (%) | Examiner fail rate (%) |
|---|---|---|---|
| Originals | 1,285 | 73.7 | 71.7 |
| Renewals | 806 | 72.1 | 62.8 |

*Note*. Results are only for test sheets that had a test score or pass/fail result recorded by an examiner.

*Test Form Difficulty and Reliability*

Table 3 shows the fail rate, mean number of errors, and internal-consistency reliability coefficient for each test form completed by each applicant group on the first test attempt. The fail rates and reliability coefficients are plotted in Figures 1 and 2, respectively.

One caveat is that some of the test forms may have been marked with a lower attempt number than was actually the case. This would have occurred if the examiner had recorded the number of the test attempt on the current paid application rather than the attempt number over all paid applications. For example, a test marked as a first attempt may actually have been an applicant's fourth try on the test (three on the first

application and one on the second application). Similarly, a test marked as a second attempt may actually have been an applicant's fifth try (three on the first application and two on the second application). The extent to which this occurred and the net biasing influence, if any, on the fail rates is unknown.

The test form fail rates for original applicants range from 68.7% to 75.8%, and average 73.4% overall. The results of the ANOVA indicate that these rates are not significantly different from one another ($F = 1.12$, $p = .35$). The test form means for number of errors made by this applicant group range from 9.7 to 10.9, and average 10.3 overall. The differences between these means approach, but do not quite reach, statistical significance ($F = 2.33$, $p = .05$). All internal-consistency reliability coefficients are above the .70 whole test minimum reliability standard, with values ranging from .77 to .84.

Table 3

Number of Tests (*n*), Fail Rate, Mean Number of Errors, and Internal-Consistency Reliability Coefficient for Each Form of the Spanish DL 5 (Rev. 6/07) on the First Test Attempt by Applicant Group

| Applicant group<br>Test form | *n* | Fail rate (%) | Mean errors | Reliability |
|---|---|---|---|---|
| Originals[a] | | | | |
| 1 | 268 | 68.7 | 9.9 | .81 |
| 2 | 273 | 74.0 | 9.7 | .77 |
| 3 | 271 | 73.4 | 10.5 | .80 |
| 4 | 257 | 75.5 | 10.9 | .84 |
| 5 | 260 | 75.8 | 10.7 | .82 |
| Total | 1,329 | 73.4 | 10.3 | .81 |
| Renewals[b] | | | | |
| 1 | 181 | 75.7 | 5.2 | .48 |
| 2 | 159 | 79.3 | 5.8 | .62 |
| 3 | 166 | 74.7 | 6.0 | .71 |
| 4 | 182 | 54.4 | 4.7 | .72 |
| 5 | 176 | 75.6 | 5.8 | .69 |
| Total | 864 | 71.6 | 5.5 | .64 |

*Note*. The figures presented for total fail rate, total mean errors, and total reliability are weighted averages. All ANOVAs are two-tailed.

[a]Forms do not differ significantly on fail rate ($F = 1.12$, $p = .35$) or mean number of errors ($F = 2.33$, $p = .05$). [b]Forms differ significantly on fail rate ($F = 9.00$, $p = .000$) and mean number of errors ($F = 5.91$, $p = .000$).

The test form fail rates for renewals range from 54.4% to 79.3%, and average 71.6% overall. The results of the omnibus ANOVA indicate that significant differences exist between these rates ($F$ = 9.00, $p$ = .000). The post hoc Games and Howell Multiple Comparisons Test identified Form 4 as being significantly different from each of the remaining forms on this measure. The test form means for number of errors range from 4.7 to 6.0, and average 5.5 overall. The differences between these means are statistically significant ($F$ = 5.91, $p$ = .000). The post hoc significance test identified Form 4 as being significantly different from Forms 2, 3, and 5 on this measure. The internal-consistency reliability coefficients range from .48 to .72, with three forms falling below the .70 standard of acceptability.

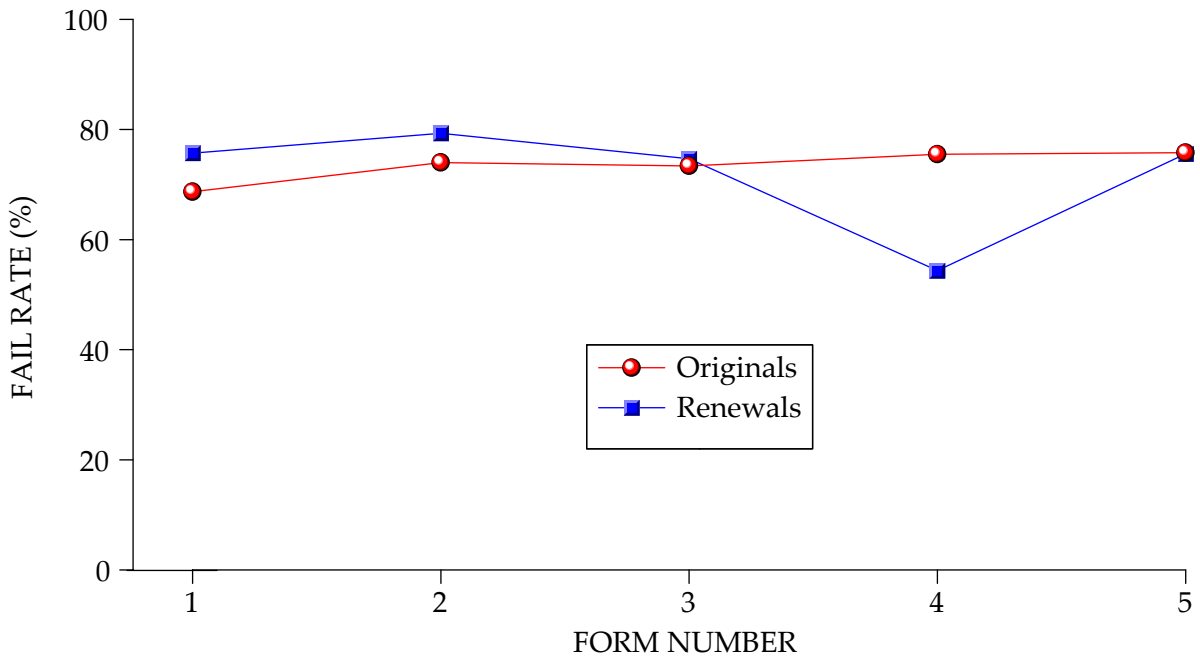


*Figure 1.* Spanish test form fail rates for original and renewal applicants on the first test attempt.

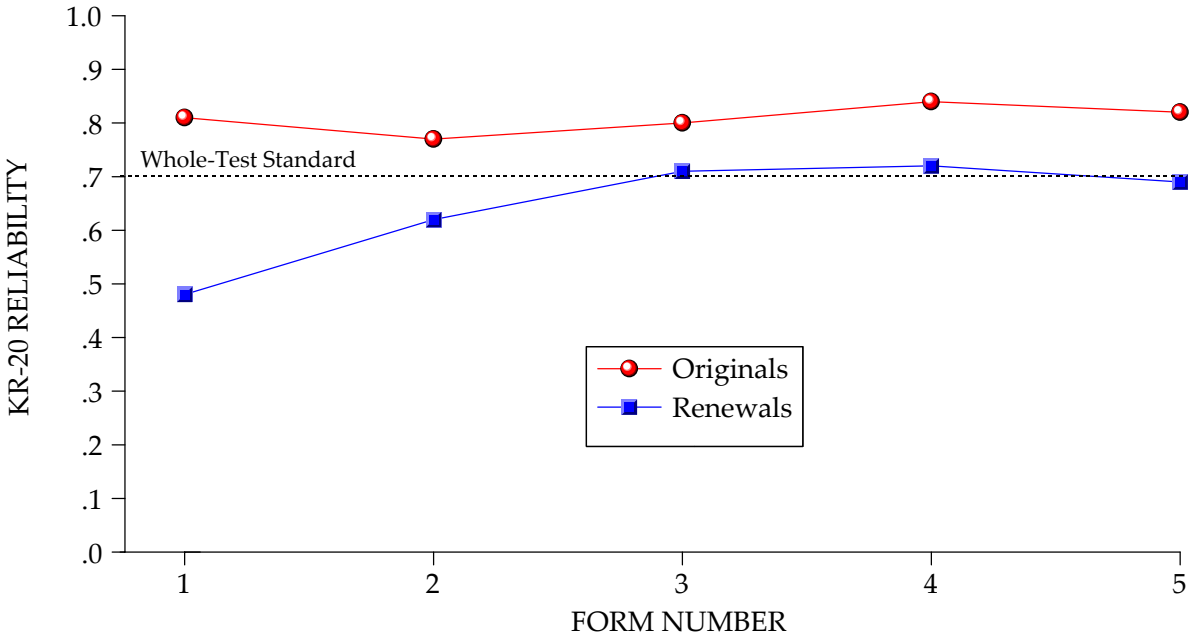


*Figure 2*. Spanish test form internal-consistency reliability coefficients for original and renewal applicants on the first test attempt.

*Test Difficulty by Attempt*

The fail rate and mean number of errors for each test attempt across all forms for each applicant group are shown in Table 4 and Figure 3. As has been found in prior statewide evaluations of the Class C written tests, the fail rate tends to remain steady or increase over the first three attempts. (The fail rates for the fourth-or-higher attempt category are not trustworthy because of the difficulty in determining attempts over multiple applications and the small sample sizes). These results likely indicate that applicants very often either did not prepare for the retests by further studying the *California Driver's Handbook*, or had problems reading or understanding the handbook or the tests.

Table 4
Number of Tests (*n*), Fail Rate, and Mean Number of Errors for the Spanish DL 5 (Rev. 6/07) For Each Test Attempt by Applicant Group

| Applicant group<br>Test attempt | *n* | Fail rate (%) | Mean errors |
|---|---|---|---|
| Originals | | | |
| First | 1,329 | 73.4 | 10.3 |
| Second | 819 | 71.3 | 9.9 |
| Third | 462 | 74.7 | 10.0 |
| Fourth or higher | 72 | 59.7 | 9.6 |
| Not Reported | 169 | 74.0 | 10.8 |
| Total | 2,851 | 72.7 | 10.2 |
| Renewals | | | |
| First | 864 | 71.6 | 5.5 |
| Second | 486 | 72.6 | 5.3 |
| Third | 210 | 78.1 | 5.8 |
| Fourth or higher | 36 | 69.4 | 5.1 |
| Not Reported | 92 | 77.2 | 5.6 |
| Total | 1,688 | 73.0 | 5.5 |

*Note*. All figures presented for total fail rate and total mean errors are weighted averages.

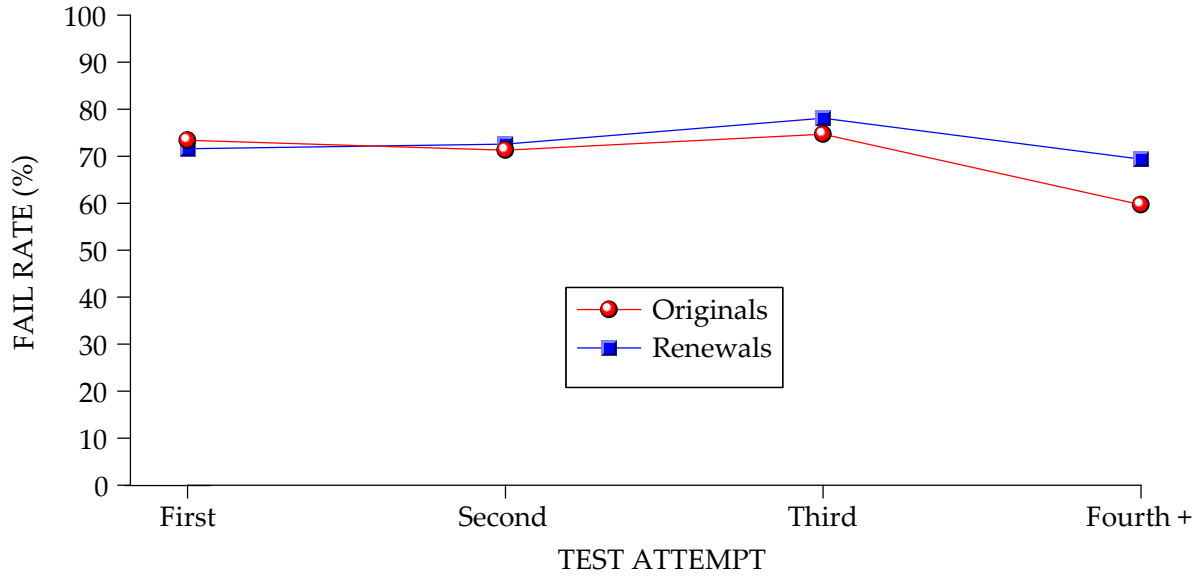


*Figure 3*. Spanish test fail rates by test attempt number for original and renewal applicants.

*Test Difficulty by Field Office*

The number of Spanish original and renewal tests received and the fail rates for these tests across all test attempts are presented for each field office in Appendix A. Fail rates computed from fewer than 20 test forms tend to lack precision and therefore may not be trustworthy. The fail rates for field offices with 20 or more forms range from 54.1% to 95.0% for original applicants and from 55.7% to 88.0% for renewal applicants. These rates indicate wide variation in the knowledge level of applicants residing in different geographical areas of the state.

## Item Statistics

All item statistics in this report are for first-attempt tests only. The results for original and renewal applicants are presented in Appendices B and C, respectively. Each of these two appendices include four tables that contain, respectively, the percentage of applicants selecting each answer choice for each item, item-total correlations, the percentage of applicants who would pass if different passing score cut-points (number wrong) were used, and a summary of potential problem items. The results presented in these tables are described in the following four subsections of this report.

*Item Pass Rates and Answer Choice Selection Rates*

The percentages of applicants selecting the different answer choices are presented in Tables B1 and C1. The underlining of a percentage indicates the correct answer choice. The results are based on valid responses only. Not represented are instances in which the applicant did not answer the item or selected more than one answer choice. The latter cases represent, for any given item, fewer than 4% of applicants taking the test.

The item pass rate refers to the percentage of applicants who answered each item correctly. Items with a pass rate lower than 60% or higher than 95% are considered to be potentially problematic items. A very low pass rate may indicate that the question is poorly worded, has ambiguous or misleading answer choices, is not adequately covered in the driver handbook, or is problematic for some other reason. A very high pass rate, on the other hand, often results when the question has one or more distracters that are so illogical or absurd that they can be easily discarded, or covers subject matter that is common knowledge and therefore doesn't need to be tested. The items with pass rates that are too high or too low are shaded in the tables.

It is desirable that the incorrect answer choices seem tenable to applicants who lack the knowledge covered by the item. Therefore, individual distracters selected by 2% or

fewer of the applicants should be considered for possible revision to increase their appeal to more applicants. Distracters selected more often than the correct answer, or within 10% of the selection rate for the correct answer, are also suspect and should be reviewed. Distracters with selection rates that are too low or too high are shown in boldface type in the table.

The statistical criteria described above are provided only as guidelines. It is not recommended that questions be revised on the basis of these statistical properties alone. In addition, in making revisions to the Spanish test questions, it is important that the subject matter content remain equivalent to that of the English tests. To accomplish this, it may be necessary to make comparable changes to the English test questions.

*Item-Total Correlations*

The item-total correlation coefficients for the test items are presented in Tables B2 and C2. Items that tended to be answered incorrectly by applicants who scored high on the test overall (i.e., with a negative coefficient value), or for which performance on the item has very little or no relationship to overall test performance (i.e., with a positive coefficient value less than .10), are highly suspect and usually (thought not in every instance) need to be modified or replaced. These items are shaded in the tables.

*Pass Rates Using Different Cut-Points*

The percentages of applicants who would pass the tests at different score cut-points are presented in Tables B3 and C3. The tables present the percentage of applicants who missed at least the number of items indicated in the leftmost column of each row, and would therefore have passed if that number of errors had been used as the minimum passing score. For example, Table B3 indicates that 25.0% of originals who completed Form 1 would have passed if the number of allowable errors had been five instead of six. The shaded row in each table shows the pass rate for each test form using the current passing standard (six for original applicants and three for renewal applicants).

*Summary of Problem Items*

Tables B4 and C4 identify items with a low item-total correlation, low or high pass rate, and/or with a distracter with a low or high selection rate, indicating that the items may need to be revised or replaced. All test forms have several items with one or more of these characteristics. Items that also appeared problematic on the English test are underlined in the table.

# DISCUSSION

The first-attempt Spanish test fail rates are lower in this evaluation than they were in the August 2005 DMV study (Reiner & Hagge, 2006). Specifically, the rate for originals decreased to 73.4% from 74.8%, and the rate for renewals decreased to 71.6% from 79.8%.

The much higher fail rates for the Spanish tests compared to those for the English tests has been a continuing concern to the department. While several possible explanations for this disparity can be hypothesized, very little research has been conducted to identify the specific causes. The Spanish tests have the same subject matter content as the English tests, as they are translated directly from the English test forms. The department conducted a thorough review of the Spanish tests a few years ago to determine whether faulty translation could possibly explain the high Spanish test fail rates, and only minor problems were found. Lack of study also does not appear to be an explanation for the higher Spanish test fail rates based on the findings in the 2006 evaluation, which showed that applicants who took the Spanish tests reported having studied the driver handbook as much as those who took the English tests. Other explanations will not be offered here because doing so without supporting research would be wholly speculative.

The results from the current evaluation show that the test forms administered to original applicants have good reliability, with all five forms having an internal-consistency reliability coefficient well above the .70 standard of acceptability. In contrast, the results show that the reliability of test forms for renewal applicants is low, with the reliability coefficients for three of the five forms falling below .70; these results have not improved since the 2006 evaluation. This continuing low reliability for the renewal test forms is not surprising, since the small number of questions on these tests makes them prone to be unreliable. This deficiency is disconcerting because it reduces the department's ability to make sound licensing decisions.

The analysis of individual test questions found that all test forms contained some items with a low item-total correlation, a pass rate that is too high or too low, or a distracter that was selected too often or too infrequently. The overall quality of the tests can be

increased by reviewing these items and then revising or replacing them as necessary. Items with low item-total correlations are especially suspect and warrant immediate attention.

Some practices followed by field office personnel were discovered over the course of this evaluation that are inconsistent with departmental policy. During the data screening process, it became apparent that some field office personnel were administering the back side rather than the front side of the DL 5 test sheet to renewal applicants. In addition, some field offices administered older revisions of the tests. These practices diminish the effectiveness of periodically randomizing the order of questions on the test forms to curtail cheating. Administering the correct side of the renewal test and using only current tests would help increase the overall integrity of the tests.

Another way to improve the integrity of the testing process would be to stop the practice of restating or substituting questions when an applicant initially misses too many questions to pass. This practice results in licensing more applicants with marginal knowledge competency. Evidence of this is provided by the substantial increase in pass rates that was found when examiner error scores rather than computer-graded error scores were used to determine test outcomes in this study. This practice undermines the department's ongoing efforts to maximize the reliability and validity of the tests through periodic evaluations and the use of state-of-the-art psychometric test construction techniques. In addition, because the practice is so subjective, it exposes the department to claims of discrimination by those who are not offered the same opportunity as others to change their original answer choices.

# RECOMMENDATIONS

There are several things the department can do to improve the quality of the Spanish written knowledge tests. In doing so, care should be taken to make similar changes to the English tests and testing requirements to maintain the equivalency of knowledge content between the two translations of the tests and to ensure that all applicant groups continue to be treated in the same manner. The following specific actions are recommended.

1. Questions with any of the following statistical characteristics should be reviewed and revised or replaced as necessary: (a) an item-total correlation coefficient lower than .10, (b) a distracter selected more often or within 10% of the correct answer, or by fewer than 2% of the respondents, or (c) a pass rate that is too high (95% or higher) or too low (60% or lower).

2. The renewal test form reliabilities should be increased so they are all at least .70. This can be accomplished by having renewal applicants complete all 36 items on the DL 5 instead of only the first 18 items. Correcting problem items, especially those with low item-total correlations, would also increase test reliability but almost certainly not to the extent necessary.

3. The order of answer choices (not just the test questions) should be periodically randomized to decrease the opportunity for cheating or rote memorization of the correct answers. Computer applications are available that can do this efficiently and cost-effectively. The department's Research and Development Branch can provide guidance in this matter if requested.

4. The department should discontinue the policy that allows examiners to restate missed test questions. This would increase the integrity of the testing process by eliminating examiner subjectivity in determining whether the customer has an adequate understanding of traffic laws and safe driving practices. It would also protect the department against charges of discrimination based on the gender, age, race, or other characteristics of the applicant rather than their knowledge competency.

5. Applicants who fail a test should be required to wait a minimum period of time before retesting. Current law requires provisional license applicants (drivers under 18 years of age) to wait at least a week between written tests (and 2 weeks between drive tests). At a minimum, applicants should not be allowed to take more than one knowledge test per day. Instituting a mandatory waiting period would increase the likelihood that applicants would study the driver handbook more thoroughly before each test attempt, which should result in improved performance.

6. Steps should be taken to ensure that field office personnel are administering only the current version of each test in accordance with the procedures stated in the department's Driver License Manual.

7. The department should continue investigating the possible use of computer technology to construct equivalent test forms from a large item-pool database. This would also support ongoing randomization of the order of test questions and answer choices.

8. The department should make the Spanish version of the driver handbook more readily available both in hardcopy and electronic form. This might include better publicizing how the handbook material can be obtained. This should also be done to the extent possible for the other language translations of the handbook.

9. The department should prepare and distribute information materials that emphasize the importance of thoroughly reading the driver handbook before taking the test. The specific knowledge content areas that are most challenging to applicants should also be publicized on the internet, in newspapers, and on television.

10. Procedures should be developed and implemented to better identify low-literacy applicants who might benefit by taking the test orally rather than in the written form.

# APPENDICES

# Appendix A

## Test Fail Rates Over All Attempts on the Spanish DL 5 (Rev. 6/07) for Original and Renewal Applicants in Each Field Office

Table A

Number of Tests (*n*) and Fail Rate Over All Attempts on the Spanish DL 5 (Rev. 6/07) for Each Applicant Group in Each Field Office

| Reporting unit number | Field office | Originals | | Renewals | |
|---|---|---|---|---|---|
| | | *n* | Fail rate (%) | *n* | Fail rate (%) |
| 501 | Sacramento | 1 | 100.0 | * | * |
| 502 | Los Angeles | 125 | 76.0 | 61 | 55.7 |
| 503 | San Francisco | 24 | 66.7 | 5 | 80.0 |
| 504 | Oakland | * | * | * | * |
| 505 | Fresno | 13 | 61.5 | 6 | 83.3 |
| 506 | San Diego | 31 | 71.0 | 10 | 80.0 |
| 507 | Long Beach | 29 | 72.4 | 9 | 77.8 |
| 508 | Hollywood | 65 | 61.5 | 37 | 78.4 |
| 509 | Pasadena | * | * | * | * |
| 510 | Glendale | 25 | 64.0 | 18 | 66.7 |
| 511 | Montebello | 61 | 72.1 | 39 | 61.5 |
| 512 | San Bernardino | 32 | 81.3 | 23 | 65.2 |
| 513 | Truckee | 4 | 75.0 | 1 | 0.0 |
| 514 | Culver City | 13 | 69.2 | 10 | 60.0 |
| 515 | Van Nuys | 38 | 76.3 | 34 | 64.7 |
| 516 | San Jose | 40 | 75.0 | 16 | 87.5 |
| 517 | Stockton | 21 | 85.7 | 12 | 83.3 |
| 518 | Mountain View | * | * | * | * |
| 519 | San Diego-Clairemont | 14 | 71.4 | 2 | 100.0 |
| 520 | Chico | 3 | 66.7 | * | * |
| 521 | Jackson | 1 | 100.0 | * | * |
| 522 | Oroville | * | * | * | * |
| 523 | Concord | 9 | 66.7 | 2 | 50.0 |
| 524 | Crescent City | * | * | * | * |
| 525 | Placerville | 4 | 100.0 | * | * |
| 526 | Eureka | 1 | 100.0 | 1 | 100.0 |
| 527 | El Centro | 67 | 76.1 | 35 | 74.3 |
| 528 | Blythe | * | * | 2 | 100.0 |
| 529 | Bakersfield | 8 | 87.5 | 4 | 100.0 |
| 530 | Lakeport | * | * | * | * |
| 531 | Susanville | * | * | * | * |
| 532 | Pomona | 34 | 70.6 | 24 | 62.5 |
| 533 | Madera | 14 | 85.7 | 8 | 75.0 |
| 534 | Corte Madera | 11 | 90.9 | 1 | 0.0 |

Table A (continued)

| Reporting unit number | Field office | Originals | | Renewals | |
|---|---|---|---|---|---|
| | | *n* | Fail rate (%) | *n* | Fail rate (%) |
| 535 | Ukiah | * | * | * | * |
| 536 | Merced | 17 | 88.2 | 9 | 66.7 |
| 537 | Alturas | * | * | * | * |
| 538 | South Lake Tahoe | 6 | 50.0 | 1 | 100.0 |
| 539 | Salinas | 32 | 65.6 | 24 | 75.0 |
| 540 | Napa | * | * | * | * |
| 541 | Grass Valley | 2 | 100.0 | 1 | 0.0 |
| 542 | Santa Ana | 74 | 71.6 | 69 | 73.9 |
| 543 | Roseville | 2 | 100.0 | 1 | 100.0 |
| 544 | Quincy | 2 | 100.0 | 1 | 0.0 |
| 545 | Riverside | 31 | 77.4 | 21 | 61.9 |
| 546 | Hollister | 4 | 50.0 | 3 | 100.0 |
| 547 | San Luis Obispo | 5 | 60.0 | * | * |
| 548 | Redwood City | 20 | 65.0 | 12 | 83.3 |
| 549 | Santa Barbara | 13 | 69.2 | 9 | 44.4 |
| 550 | Capitola | 2 | 100.0 | 5 | 100.0 |
| 551 | Redding | * | * | * | * |
| 552 | Yreka | * | * | * | * |
| 553 | Tulelake | * | * | * | * |
| 554 | Vallejo | 7 | 57.1 | 2 | 100.0 |
| 555 | Santa Rosa | 20 | 60.0 | 7 | 100.0 |
| 556 | El Cerrito | 14 | 64.3 | 9 | 77.8 |
| 557 | Modesto | 23 | 91.3 | 24 | 79.2 |
| 558 | Red Bluff | 5 | 80.0 | 1 | 100.0 |
| 559 | Visalia | 12 | 66.7 | 5 | 60.0 |
| 560 | Ventura | 6 | 66.7 | 8 | 87.5 |
| 561 | Woodland | 20 | 65.0 | 12 | 58.3 |
| 562 | Yuba City | 9 | 88.9 | 5 | 60.0 |
| 563 | Santa Maria | 11 | 90.9 | 8 | 75.0 |
| 564 | Colusa | 10 | 70.0 | 4 | 100.0 |
| 565 | Hanford | 9 | 66.7 | 12 | 83.3 |
| 566 | Mariposa | * | * | 2 | 100.0 |
| 567 | Seaside | 3 | 0.0 | 3 | 66.7 |
| 568 | San Andreas | * | * | * | * |
| 569 | Sonora | * | * | * | * |
| 570 | Auburn | * | * | 1 | 0.0 |
| 571 | Willows | * | * | * | * |

Table A (continued)

| Reporting unit number | Field office | Originals | | Renewals | |
|---|---|---|---|---|---|
| | | *n* | Fail rate (%) | *n* | Fail rate (%) |
| 572 | Weaverville | * | * | * | * |
| 573 | Porterville | 23 | 73.9 | 10 | 80.0 |
| 574 | Paso Robles | 5 | 80.0 | 2 | 100.0 |
| 575 | Taft | * | * | * | * |
| 576 | Bell Gardens | 115 | 70.4 | 80 | 66.3 |
| 577 | Ridgecrest | 1 | 100.0 | 1 | 0.0 |
| 578 | Indio | 61 | 75.4 | 25 | 88.0 |
| 579 | Hayward | 29 | 79.3 | 9 | 77.8 |
| 580 | Clovis | 10 | 80.0 | 3 | 100.0 |
| 581 | Compton | 43 | 79.1 | 41 | 65.9 |
| 582 | Barstow | 1 | 100.0 | 3 | 100.0 |
| 583 | Watsonville | 21 | 71.4 | 14 | 50.0 |
| 584 | Needles | * | * | 1 | 0.0 |
| 585 | Bishop | 18 | 72.2 | 15 | 60.0 |
| 586 | Norco | 14 | 85.7 | 16 | 75.0 |
| 587 | Arleta | 73 | 74.0 | 53 | 77.4 |
| 588 | Vacaville | * | * | * | * |
| 589 | Lompoc | 3 | 66.7 | 3 | 100.0 |
| 590 | Fort Bragg | 2 | 100.0 | 2 | 50.0 |
| 591 | Whittier | 37 | 86.5 | 32 | 71.9 |
| 592 | Pittsburg | 17 | 82.4 | 5 | 80.0 |
| 593 | San Mateo | 22 | 77.3 | 2 | 100.0 |
| 594 | Tulare | 8 | 87.5 | 2 | 100.0 |
| 595 | Lancaster | 10 | 90.0 | 2 | 50.0 |
| 596 | Oceanside | 45 | 82.2 | 18 | 77.8 |
| 597 | Brawley | 33 | 69.7 | 15 | 80.0 |
| 598 | Davis | * | * | * | * |
| 599 | Daly City | 19 | 78.9 | 10 | 100.0 |
| 601 | Paradise | * | * | * | * |
| 602 | Sacramento-South | * | * | * | * |
| 603 | Coalinga | 11 | 54.5 | 4 | 75.0 |
| 604 | Oakland Coliseum | 20 | 95.0 | 11 | 81.8 |
| 605 | Laguna Hills | 8 | 50.0 | 5 | 60.0 |
| 606 | Bellflower | 33 | 66.7 | 28 | 75.0 |
| 607 | Fullerton | 49 | 75.5 | 39 | 69.2 |
| 608 | Torrance | 18 | 72.2 | 16 | 75.0 |
| 609 | Hawthorne | 54 | 77.8 | 27 | 81.5 |

Table A (continued)

| Reporting unit number | Field office | Originals | | Renewals | |
|---|---|---|---|---|---|
| | | *n* | Fail rate (%) | *n* | Fail rate (%) |
| 610 | Inglewood | 13 | 53.8 | 15 | 93.3 |
| 611 | Westminster | 37 | 54.1 | 12 | 58.3 |
| 612 | Rancho Cucamonga | 29 | 58.6 | 18 | 72.2 |
| 613 | Chula Vista | 80 | 71.3 | 40 | 85.0 |
| 614 | Spring Valley | 1 | 0.0 | 2 | 50.0 |
| 615 | Delano | * | * | * | * |
| 616 | Santa Monica | 9 | 77.8 | 5 | 80.0 |
| 617 | Lincoln Park | 53 | 69.8 | 37 | 78.4 |
| 618 | West Covina | 46 | 54.3 | 45 | 75.6 |
| 619 | San Pedro | 22 | 77.3 | 20 | 85.0 |
| 620 | Escondido | 32 | 84.4 | 15 | 73.3 |
| 621 | Fairfield | * | * | * | * |
| 622 | Lodi | 19 | 73.7 | 7 | 85.7 |
| 623 | Gilroy | 5 | 100.0 | 5 | 60.0 |
| 624 | Walnut Creek | 10 | 80.0 | 2 | 0.0 |
| 625 | Carmichael | * | * | * | * |
| 626 | Redlands | 15 | 73.3 | 7 | 57.1 |
| 627 | Garberville | * | * | * | * |
| 628 | Costa Mesa | 8 | 75.0 | 2 | 50.0 |
| 629 | Victorville | 40 | 65.0 | 10 | 40.0 |
| 630 | Santa Paula | 9 | 66.7 | 8 | 62.5 |
| 631 | Pleasanton | * | * | * | * |
| 632 | Santa Clara | 10 | 80.0 | 6 | 83.3 |
| 633 | Reedley | 21 | 76.2 | 26 | 84.6 |
| 634 | Petaluma | 8 | 62.5 | 2 | 100.0 |
| 635 | Hemet | 27 | 77.8 | 18 | 72.2 |
| 636 | Oxnard | 58 | 74.1 | 33 | 87.9 |
| 637 | Winnetka | 14 | 85.7 | 6 | 100.0 |
| 638 | Twentynine Palms | 1 | 0.0 | * | * |
| 639 | Mount Shasta | * | * | 2 | 50.0 |
| 640 | Los Gatos | 5 | 100.0 | 5 | 80.0 |
| 641 | Banning | 5 | 60.0 | 4 | 100.0 |
| 642 | Tracy | 12 | 66.7 | 4 | 100.0 |
| 643 | Fall River Mills | * | * | * | * |
| 644 | Fremont | 18 | 72.2 | 7 | 71.4 |
| 645 | Orland | * | * | * | * |
| 646 | Fresno-North | 5 | 60.0 | 7 | 57.1 |

Table A (continued)

| Reporting unit number | Field office | Originals | | Renewals | |
|---|---|---|---|---|---|
| | | *n* | Fail rate (%) | *n* | Fail rate (%) |
| 647 | King City | 6 | 83.3 | 2 | 100.0 |
| 648 | San Clemente | 3 | 66.7 | 6 | 33.3 |
| 649 | Turlock | 11 | 63.6 | 5 | 100.0 |
| 650 | Los Banos | 11 | 72.7 | 13 | 69.2 |
| 655 | Folsom | * | * | * | * |
| 656 | Riverside-East | 26 | 76.9 | 18 | 72.2 |
| 657 | Fontana | 45 | 75.6 | 36 | 72.2 |
| 658 | Manteca | 14 | 85.7 | 10 | 80.0 |
| 659 | Palm Springs | * | * | * | * |
| 660 | Shafter | 11 | 72.7 | 13 | 69.2 |
| 661 | Arvin | 14 | 71.4 | 5 | 60.0 |
| 662 | Newhall | 7 | 42.9 | 9 | 55.6 |
| 663 | Thousand Oaks | 6 | 83.3 | 4 | 75.0 |
| 668 | Santa Teresa | 13 | 76.9 | 8 | 62.5 |
| 669 | El Cajon | 20 | 70.0 | 5 | 60.0 |
| 670 | Goleta | 3 | 100.0 | 2 | 50.0 |
| 672 | Temecula | 32 | 81.3 | 9 | 66.7 |
| 673 | Rocklin | 1 | 100.0 | * | * |
| 675 | Poway Center (PWC) | 2 | 50.0 | 2 | 100.0 |
| 676 | Poway | * | * | * | * |
| 677 | San Ysidro | 112 | 58.0 | 42 | 69.0 |
| 679 | Bakersfield-Southwest | 8 | 75.0 | 3 | 100.0 |
| 680 | Simi Valley | 10 | 70.0 | 2 | 50.0 |
| 686 | Novato | 12 | 66.7 | 1 | 100.0 |
| 687 | Lake Isabella | * | * | * | * |
| 690 | Palmdale | 21 | 81.0 | 9 | 88.9 |
| 691 | Placentia | * | * | * | * |
| 697 | West Sacramento | * | * | * | * |
| 699 | San Bernardino | * | * | * | * |

*Note.* Fail rates based on fewer than 20 test sheets are likely to be unreliable and should not be interpreted as accurate. An asterisk (*) entry indicates that no test forms were received.

# Appendix B

## Item Statistics for Original Applicants on the Spanish DL 5 (Rev. 6/07) First Test Attempt

Table B1

Percentage of Original Applicants Selecting Each Answer Choice for Each Item on Each Form of the Spanish DL 5 (Rev. 6/07) on the First Test Attempt

| Item | Answer choice | Form 1 (*n* = 268) | Form 2 (*n* = 273) | Form 3 (*n* = 271) | Form 4 (*n* = 257) | Form 5 (*n* = 260) |
|---|---|---|---|---|---|---|
| 1 | a | 6.7 | 72.7 | 92.3 | 8.6 | 8.1 |
| | b | 4.9 | 8.1 | 4.1 | 73.0 | **1.5** |
| | c | 88.4 | 19.2 | 3.7 | 18.4 | 90.3 |
| 2 | a | 74.1 | 93.4 | 79.9 | 8.7 | 6.6 |
| | b | 6.8 | 3.3 | 6.3 | 88.5 | 76.0 |
| | c | 19.2 | 3.3 | 13.8 | 2.8 | 17.4 |
| 3 | a | 13.1 | 19.1 | 16.4 | 4.3 | 20.2 |
| | b | 3.7 | 68.5 | 55.2 | 80.6 | 7.8 |
| | c | 83.2 | 12.4 | 28.4 | 15.0 | 72.0 |
| 4 | a | 79.2 | 20.0 | 15.5 | **2.0** | 27.1 |
| | b | 13.3 | 20.4 | 15.9 | **2.0** | 63.6 |
| | c | 7.6 | 59.6 | 68.6 | 96.1 | 9.3 |
| 5 | a | 7.1 | 37.5 | 8.9 | 82.3 | 15.1 |
| | b | 7.8 | **57.7** | 68.3 | 3.5 | 19.8 |
| | c | 85.1 | 4.9 | 22.9 | 14.2 | 65.1 |
| 6 | a | 9.9 | 9.7 | 80.0 | 70.7 | 72.0 |
| | b | 75.7 | 78.4 | 9.6 | 15.2 | 20.1 |
| | c | 14.4 | 11.9 | 10.4 | 14.1 | 7.9 |
| 7 | a | 21.1 | 6.3 | 85.2 | 15.6 | 21.2 |
| | b | 72.5 | 72.8 | 5.9 | 10.5 | 7.3 |
| | c | 6.4 | 20.9 | 8.9 | 73.9 | 71.5 |
| 8 | a | 27.0 | 10.6 | **45.4** | 5.6 | 28.9 |
| | b | 66.5 | 4.4 | 37.5 | 90.8 | 67.6 |
| | c | 6.5 | 85.0 | 17.1 | 3.6 | 3.5 |
| 9 | a | 20.6 | 80.1 | 3.0 | **1.6** | 21.2 |
| | b | 71.2 | 9.6 | **56.3** | 83.4 | 19.3 |
| | c | 8.2 | 10.3 | 40.7 | 15.0 | 59.5 |
| 10 | a | 94.0 | 62.5 | 21.5 | 76.6 | 10.9 |
| | b | 3.0 | 20.4 | 6.7 | 4.7 | 80.1 |
| | c | 3.0 | 17.1 | 71.9 | 18.8 | 9.0 |
| 11 | a | 9.5 | 36.4 | 82.6 | 7.8 | 46.5 |
| | b | 9.9 | 4.0 | 7.0 | 18.0 | 9.4 |
| | c | 80.5 | **59.6** | 10.4 | 74.1 | **44.1** |

Table B1 (continued)

| Item | Answer choice | Form 1 ($n$ = 268) | Form 2 ($n$ = 273) | Form 3 ($n$ = 271) | Form 4 ($n$ = 257) | Form 5 ($n$ = 260) |
|---|---|---|---|---|---|---|
| 12 | a | 56.9 | 2.6 | 7.5 | 10.6 | 65.3 |
| | b | 29.2 | 88.6 | 11.6 | 56.7 | 17.4 |
| | c | 13.9 | 8.8 | 81.0 | 32.7 | 17.4 |
| 13 | a | 4.6 | 80.3 | 31.6 | 82.7 | 76.9 |
| | b | 53.1 | 13.0 | 48.3 | 9.4 | 17.7 |
| | c | 42.3 | 6.7 | 20.2 | 7.8 | 5.4 |
| 14 | a | 9.7 | 18.0 | 9.6 | 5.2 | 3.9 |
| | b | 86.6 | 26.1 | 78.9 | 4.8 | 63.3 |
| | c | 3.7 | 55.9 | 11.5 | 90.1 | 32.8 |
| 15 | a | 13.6 | 14.3 | 28.8 | 77.6 | 17.3 |
| | b | 74.3 | 82.4 | 66.1 | 12.2 | 77.3 |
| | c | 12.1 | 3.3 | 5.2 | 10.2 | 5.4 |
| 16 | a | 16.9 | 10.7 | 46.5 | 10.4 | 81.5 |
| | b | 10.1 | 14.0 | **40.1** | 71.1 | 7.7 |
| | c | 73.0 | 75.4 | 13.4 | 18.5 | 10.8 |
| 17 | a | 2.7 | 11.1 | 7.2 | 20.5 | 18.9 |
| | b | 12.5 | 3.3 | 83.7 | 53.9 | 61.4 |
| | c | 84.8 | 85.6 | 9.1 | 25.6 | 19.7 |
| 18 | a | 13.1 | 97.1 | 84.0 | 3.5 | 23.3 |
| | b | 31.7 | **1.8** | 7.1 | 73.4 | 14.4 |
| | c | 55.2 | **1.1** | 8.9 | 23.0 | 62.3 |
| 19 | a | 78.9 | 78.2 | 3.0 | 57.3 | 57.2 |
| | b | 5.7 | 10.3 | 3.0 | 29.4 | 13.5 |
| | c | 15.5 | 11.4 | 94.1 | 13.3 | 19.3 |
| 20 | a | 79.3 | 3.7 | 14.1 | 71.9 | 34.6 |
| | b | 13.2 | 15.4 | 79.3 | 15.2 | 18.8 |
| | c | 7.5 | 80.9 | 6.7 | 12.9 | **46.5** |
| 21 | a | 14.3 | 33.0 | 6.7 | 13.7 | 8.6 |
| | b | 66.4 | 62.2 | 16.0 | 63.9 | 86.3 |
| | c | 19.2 | 4.8 | 77.2 | 22.5 | 5.1 |
| 22 | a | 8.2 | 83.7 | 7.8 | 15.7 | 11.2 |
| | b | 7.5 | 8.5 | 77.2 | 66.9 | 36.8 |
| | c | 84.3 | 7.8 | 14.9 | 17.3 | 51.9 |
| 23 | a | 53.8 | 6.6 | 6.8 | 15.8 | 18.9 |
| | b | 19.8 | 5.2 | 42.1 | 62.1 | 10.0 |
| | c | 26.3 | 88.2 | **51.1** | 22.1 | 71.0 |

Table B1 (continued)

| Item | Answer choice | Form 1 (*n* = 268) | Form 2 (*n* = 273) | Form 3 (*n* = 271) | Form 4 (*n* = 257) | Form 5 (*n* = 260) |
|---|---|---|---|---|---|---|
| 24 | a | 67.8 | 19.7 | 2.2 | 55.9 | 6.6 |
| | b | 25.5 | 7.8 | 4.8 | 24.8 | 30.5 |
| | c | 6.7 | 72.5 | 92.9 | 19.3 | 62.9 |
| 25 | a | 16.0 | 78.3 | **43.3** | 33.7 | 50.8 |
| | b | 61.6 | 14.7 | 15.7 | 50.0 | 15.1 |
| | c | 22.4 | 7.0 | 41.0 | 16.3 | 34.1 |
| 26 | a | 93.6 | 8.1 | 17.8 | 9.8 | 3.1 |
| | b | 5.7 | 10.7 | 73.6 | 11.4 | 91.9 |
| | c | **.8** | 81.2 | 8.6 | 78.8 | 5.0 |
| 27 | a | 78.5 | 4.8 | 8.9 | 2.4 | 3.5 |
| | b | 14.7 | 71.5 | 22.7 | 92.0 | 10.0 |
| | c | 6.8 | 23.7 | 68.4 | 5.6 | 86.5 |
| 28 | a | 8.8 | 78.4 | 9.0 | 34.0 | 4.6 |
| | b | 22.9 | 15.2 | 15.0 | 8.2 | **1.5** |
| | c | 68.3 | 6.4 | 76.0 | 57.8 | 93.8 |
| 29 | a | 67.8 | 71.5 | 11.3 | 9.7 | 6.2 |
| | b | 18.9 | 15.2 | 79.0 | 61.5 | 41.7 |
| | c | 13.3 | 13.3 | 9.4 | 28.8 | 52.1 |
| 30 | a | 19.2 | 57.4 | 20.2 | 59.3 | 24.3 |
| | b | 50.0 | 5.2 | 55.5 | 20.6 | 64.9 |
| | c | 30.8 | 37.4 | 24.3 | 20.2 | 10.8 |
| 31 | a | 17.6 | 6.3 | 70.3 | 52.4 | 79.8 |
| | b | 13.5 | 42.6 | 17.5 | 40.9 | 17.4 |
| | c | 68.9 | **51.1** | 12.3 | 6.7 | 2.7 |
| 32 | a | 53.4 | 12.7 | 18.7 | 10.1 | 19.6 |
| | b | 34.2 | 14.2 | 8.2 | 42.0 | 78.8 |
| | c | 12.4 | 73.1 | 73.1 | **47.9** | **1.5** |
| 33 | a | 7.5 | **1.5** | **1.2** | **1.2** | 24.0 |
| | b | 9.3 | 94.5 | 15.0 | 9.8 | 4.7 |
| | c | 83.2 | 4.0 | 83.8 | 89.1 | 71.3 |
| 34 | a | 8.2 | 76.8 | 84.8 | 36.5 | 5.1 |
| | b | 88.8 | 18.4 | 12.6 | 10.2 | 9.7 |
| | c | 3.0 | 4.9 | 2.6 | **53.3** | 85.2 |
| 35 | a | 11.3 | 6.6 | 13.4 | 80.9 | 8.1 |
| | b | 66.8 | 10.3 | 2.2 | 16.0 | 76.4 |
| | c | 21.9 | 93.1 | 84.3 | 3.1 | 15.5 |

Table B1 (continued)

| Item | Answer choice | Form 1 ($n$ = 268) | Form 2 ($n$ = 273) | Form 3 ($n$ = 271) | Form 4 ($n$ = 257) | Form 5 ($n$ = 260) |
|---|---|---|---|---|---|---|
| 36 | a | 74.3 | 8.1 | 9.3 | 20.0 | 88.0 |
| | b | 15.8 | 66.9 | 75.4 | 64.7 | 5.8 |
| | c | 9.8 | 25.0 | 15.3 | 15.3 | 6.2 |

*Note*. The entry for $n$ at the top of each column is the number of usable first-attempt test sheets that were used to compute the item choice response rates in the column. The percentages are based on responses in which one, and only one, answer choice was selected for the particular item. Underlining of a percentage indicates that the answer choice was the correct response. Shading indicates that the item has a pass rate that is too low (< 60%) or too high (> 95%) and therefore needs to be reviewed and possibly revised or replaced. A boldface percentage indicates that the distracter selection rate is too low (≤ 2%) or too high (within 10% or higher than the correct answer) and therefore the distracter may need to be revised or replaced. Item choice selection rates may not add to 100.0% due to rounding.

Table B2

Item-Total Correlation Coefficient for Each Item on Each Form of the Spanish DL 5 (Rev. 6/07) for Original Applicants on the First Test Attempt

| Item | Form 1 ($n$ = 268) | Form 2 ($n$ = 273) | Form 3 ($n$ = 271) | Form 4 ($n$ = 257) | Form 5 ($n$ = 260) |
|---|---|---|---|---|---|
| 1 | .29 | .15 | .22 | .34 | .18 |
| 2 | .28 | .14 | .22 | .22 | .20 |
| 3 | .20 | .32 | .19 | .25 | .39 |
| 4 | .23 | .09 | .33 | .25 | .21 |
| 5 | .26 | .36 | .33 | .27 | .46 |
| 6 | .35 | .39 | .29 | .36 | .43 |
| 7 | .39 | .36 | .24 | .46 | .28 |
| 8 | .42 | .35 | .39 | .30 | .16 |
| 9 | .18 | .19 | .38 | .22 | .24 |
| 10 | .41 | .24 | .28 | .39 | .12 |
| 11 | .28 | .09 | .18 | .20 | .27 |
| 12 | .26 | .40 | .22 | .45 | .34 |
| 13 | .51 | .19 | .31 | .34 | .25 |
| 14 | .27 | .32 | .31 | .29 | .39 |
| 15 | .39 | .20 | .37 | .35 | .32 |
| 16 | .14 | .35 | .41 | .26 | .45 |
| 17 | .30 | .21 | .22 | .38 | .24 |
| 18 | .32 | .03 | .44 | .37 | .22 |
| 19 | .24 | .19 | .23 | .31 | .34 |
| 20 | .25 | .28 | .22 | .37 | .35 |
| 21 | .24 | .23 | .31 | .30 | .34 |
| 22 | .43 | .31 | .35 | .16 | .43 |
| 23 | .46 | .24 | .32 | .33 | .32 |
| 24 | .24 | .24 | .25 | .36 | .39 |
| 25 | .35 | .07 | .10 | .35 | .46 |
| 26 | .41 | .26 | .37 | .33 | .22 |
| 27 | .29 | .29 | .38 | .33 | .13 |

Table B2 (continued)

| Item | Form 1 ($n$ = 268) | Form 2 ($n$ = 273) | Form 3 ($n$ = 271) | Form 4 ($n$ = 257) | Form 5 ($n$ = 260) |
|---|---|---|---|---|---|
| 28 | .19 | .17 | .21 | .28 | .24 |
| 29 | .30 | .34 | .31 | .30 | .30 |
| 30 | .39 | .39 | .17 | .28 | .36 |
| 31 | .43 | .31 | .24 | .43 | .09 |
| 32 | .14 | .44 | .36 | .38 | .45 |
| 33 | .06 | .22 | .40 | .35 | .32 |
| 34 | .23 | .25 | .19 | .20 | .30 |
| 35 | .24 | .31 | .17 | .40 | .35 |
| 36 | .35 | .31 | .36 | .46 | .23 |

*Note*. The entry for $n$ at the top of each column is the number of usable first-attempt tests that were used to compute the item-total correlations in the column. Shading indicates that the item has a item-total correlation that is negative or less than .10 and therefore needs to be reviewed and possibly revised or replaced.

Table B3

Percentage of Original Applicants Who Would Have Passed on the First Test Attempt if Different Passing Cut-Points (Number Wrong) Had Been Used for Each Form of the Spanish DL 5 (Rev. 6/07)

| Number wrong | Form 1 ($n$ = 268) | Form 2 ($n$ = 273) | Form 3 ($n$ = 271) | Form 4 ($n$ = 257) | Form 5 ($n$ = 260) | Total ($N$ = 1,329) |
|---|---|---|---|---|---|---|
| 0 | 3.7 | 1.5 | 2.2 | .8 | 2.7 | 2.2 |
| 1 | 5.6 | 2.9 | 4.8 | 5.1 | 5.8 | 4.8 |
| 2 | 10.4 | 5.9 | 6.3 | 11.7 | 9.2 | 8.7 |
| 3 | 13.8 | 11.0 | 12.5 | 14.4 | 13.8 | 13.1 |
| 4 | 21.6 | 17.6 | 14.8 | 18.7 | 15.4 | 17.6 |
| 5 | 25.0 | 21.6 | 19.6 | 21.0 | 18.8 | 21.2 |
| 6 | 31.3 | 26.0 | 26.6 | 24.5 | 24.2 | 26.6 |
| 7 | 36.9 | 34.4 | 31.0 | 28.8 | 30.4 | 32.4 |
| 8 | 41.0 | 43.6 | 34.7 | 36.2 | 35.4 | 38.2 |
| 9 | 44.4 | 49.1 | 42.4 | 40.1 | 39.2 | 43.1 |
| 10 | 51.9 | 58.6 | 50.6 | 49.8 | 50.4 | 52.3 |
| 11 | 60.8 | 65.6 | 56.5 | 55.6 | 56.2 | 59.0 |
| 12 | 68.3 | 74.0 | 65.7 | 60.3 | 65.0 | 66.7 |
| 13 | 74.6 | 79.1 | 73.4 | 66.5 | 70.0 | 72.8 |
| 14 | 81.3 | 85.0 | 79.0 | 72.0 | 75.8 | 78.7 |
| 15 | 85.8 | 89.0 | 83.0 | 77.4 | 78.8 | 82.9 |
| 16+ | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

*Note*. The entry for $n$ at the top of each column is the number of usable tests that were used to compute the percentages in the column. The shading highlights the pass rates at the current passing standard of six allowable errors.

Table B4

Identification of Problem Items on Each Form of the Spanish DL 5 (Rev. 6/07) Test for Original Applicants on the First Test Attempt

| Problem indicator | Form 1 Items | Form 2 Items | Form 3 Items | Form 4 Items | Form 5 Items |
|---|---|---|---|---|---|
| Pass rate too high[a] | | 18 | | 4 | |
| Pass rate too low[b] | 12, 13, 18, 23, 30, 32 | 4, 5, 11, 14, 30, 31 | 3, 8, 9, 13, 16, 23, 25, 30 | 12, 17, 19, 24, 25, 28, 30, 31, 32, 34 | 9, 11, 19, 20, 22, 25, 29 |
| Item-total correlation too low or negative[c] | 33 | 4, 11, 18, 25 | | | 31 |
| Distracter selected too often[d] | | 5, 11, 31 | 8, 9, 16, 23, 25 | 32, 34 | 11, 20 |
| Distracter selected too infrequently[e] | 26 | 18, 33 | 33 | 4, 9, 33 | 1, 28, 32 |

*Note*. The entries in the table are the test question numbers printed on the test form. An item may be represented in more than one problem category. The underlining of an item indicates that the item also appeared to be problematic on the English test translation.

[a]More than 95% of applicants answered the item correctly. [b]Less than 60% of applicants answered the item correctly. [c]The item-total correlation was negative or less than .10. [d]The distracter was chosen more often, or almost as often, as the correct answer. [e]A distracter was selected by 2% or fewer applicants.

# Appendix C

# Item Statistics for Renewal Applicants on the Spanish DL 5 (Rev. 6/07) First Test Attempt

Table C1

Percentage of Renewal Applicants Selecting Each Answer Choice for Each Item on Each Form of the Spanish DL 5 (Rev. 6/07) on the First Test Attempt

| Item | Answer choice | Form 1 ($n$ = 181) | Form 2 ($n$ = 159) | Form 3 ($n$ = 166) | Form 4 ($n$ = 182) | Form 5 ($n$ = 176) |
|---|---|---|---|---|---|---|
| 1 | a | 6.7 | 75.8 | 94.6 | 6.1 | 11.4 |
| | b | 2.8 | 7.0 | 2.4 | 68.5 | **1.7** |
| | c | 90.4 | 17.2 | 3.0 | 25.4 | 86.9 |
| 2 | a | 71.1 | 96.9 | 71.3 | 17.8 | 8.0 |
| | b | 7.2 | **1.9** | 4.4 | 79.4 | 71.6 |
| | c | 21.7 | **1.3** | 24.4 | 2.8 | 20.5 |
| 3 | a | 23.5 | 26.5 | 29.9 | 4.4 | 12.1 |
| | b | 11.7 | 52.3 | 40.9 | 83.4 | 10.3 |
| | c | 64.8 | 21.3 | 29.3 | 12.2 | 77.6 |
| 4 | a | 80.6 | 9.0 | 12.7 | 4.4 | 27.0 |
| | b | 9.4 | 18.1 | 12.1 | **1.7** | 62.6 |
| | c | 10.0 | 72.9 | 75.2 | 93.9 | 10.3 |
| 5 | a | 7.2 | 20.6 | 9.4 | 79.8 | 21.1 |
| | b | 6.6 | **77.4** | 65.6 | **1.1** | 22.3 |
| | c | 86.2 | **1.9** | 25.0 | 19.1 | 56.6 |
| 6 | a | 7.3 | 17.4 | 73.6 | 68.5 | 64.9 |
| | b | 68.0 | 59.4 | 10.4 | 12.2 | 24.1 |
| | c | 24.7 | 23.2 | 16.0 | 19.3 | 10.9 |
| 7 | a | 30.6 | 6.9 | 89.2 | 16.0 | 21.6 |
| | b | 61.7 | 66.7 | 4.2 | 12.7 | 3.4 |
| | c | 7.8 | 26.4 | 6.6 | 71.3 | 75.0 |
| 8 | a | 22.2 | 9.5 | **49.1** | 5.6 | 27.8 |
| | b | 71.7 | 6.3 | 36.0 | 90.6 | 68.2 |
| | c | 6.1 | 84.2 | 14.9 | 3.9 | 4.0 |
| 9 | a | 24.0 | 77.7 | 3.7 | 3.3 | 24.3 |
| | b | 71.5 | 8.3 | **62.3** | 79.6 | 14.5 |
| | c | 4.5 | 14.0 | 34.0 | 17.1 | 61.3 |
| 10 | a | 90.6 | 50.0 | 22.7 | 65.2 | 13.2 |
| | b | 5.5 | 34.0 | 8.0 | 7.9 | 77.6 |
| | c | 3.9 | 16.0 | 69.3 | 27.0 | 9.2 |
| 11 | a | 2.8 | 26.8 | 66.9 | 7.2 | 40.1 |
| | b | 7.8 | 4.5 | 14.5 | 30.4 | 6.4 |
| | c | 89.4 | **68.8** | 18.7 | 62.4 | **53.5** |

Table C1 (continued)

| Item | Answer choice | Form 1 ($n$ = 181) | Form 2 ($n$ = 159) | Form 3 ($n$ = 166) | Form 4 ($n$ = 182) | Form 5 ($n$ = 176) |
|---|---|---|---|---|---|---|
| 12 | a | 57.5 | **1.9** | 2.4 | 12.4 | 73.9 |
| | b | 32.6 | 83.5 | 12.1 | 62.9 | 11.9 |
| | c | 9.9 | 14.6 | 85.5 | 24.7 | 14.2 |
| 13 | a | 3.4 | 65.0 | 38.4 | 78.5 | 84.7 |
| | b | 37.4 | 25.5 | 50.6 | 9.4 | 5.7 |
| | c | **59.2** | 9.6 | 11.0 | 12.2 | 9.7 |
| 14 | a | 12.4 | 18.4 | 9.1 | 5.5 | 9.7 |
| | b | 81.5 | 19.6 | 87.8 | 4.4 | 49.7 |
| | c | 6.2 | 62.0 | 3.0 | 90.1 | **40.6** |
| 15 | a | 18.6 | 12.6 | 26.2 | 73.3 | 30.5 |
| | b | 64.4 | 85.5 | 67.1 | 12.2 | 58.6 |
| | c | 16.9 | **1.9** | 6.7 | 14.4 | 10.9 |
| 16 | a | 23.5 | 6.4 | 45.7 | 13.6 | 75.1 |
| | b | 15.6 | 12.7 | **47.5** | 65.5 | 6.4 |
| | c | 60.9 | 80.9 | 6.8 | 20.9 | 18.5 |
| 17 | a | **0.6** | 22.6 | 3.7 | 17.2 | 8.5 |
| | b | 8.9 | **1.9** | 81.7 | 54.4 | 75.6 |
| | c | 90.5 | 75.5 | 14.6 | 28.3 | 15.9 |
| 18 | a | 10.0 | 98.1 | 80.1 | 6.0 | 23.6 |
| | b | **40.0** | **1.3** | 6.0 | 78.6 | 12.1 |
| | c | 50.0 | **0.6** | 13.9 | 15.4 | 64.4 |

*Note*. The entry for $n$ at the top of each column is the number of usable first-attempt test sheets that were used to compute the item choice response rates in the column. The percentages are based on responses in which one, and only one, answer choice was selected for the particular item. Underlining of a percentage indicates that the answer choice was the correct response. Shading indicates that the item has a pass rate that is too low (< 60%) or too high (> 95%) and therefore needs to be reviewed and possibly revised or replaced. A boldface percentage indicates that the distracter selection rate is too low (≤ 2%) or too high (within 10% or higher than the correct answer) and therefore the distracter may need to be revised or replaced. Item choice selection rates may not add to 100.0% due to rounding.

Table C2

Item-Total Correlation Coefficient for Each Item on Each Form of the Spanish DL 5 (Rev. 6/07) for Renewal Applicants on the First Test Attempt

| Item | Form 1 ($n$ = 181) | Form 2 ($n$ = 159) | Form 3 ($n$ = 166) | Form 4 ($n$ = 182) | Form 5 ($n$ = 176) |
|---|---|---|---|---|---|
| 1 | .30 | .23 | .27 | .26 | .41 |
| 2 | .19 | .15 | .33 | .40 | .26 |
| 3 | -.01 | .34 | .37 | .34 | .32 |
| 4 | .31 | .18 | .37 | .35 | .24 |
| 5 | .21 | .21 | .30 | .29 | .31 |
| 6 | .10 | .26 | .34 | .22 | .28 |
| 7 | .31 | .17 | .30 | .32 | .27 |
| 8 | .16 | .36 | .29 | .48 | .10 |
| 9 | -.26 | .26 | .26 | .13 | .29 |
| 10 | .42 | .27 | .22 | .47 | .16 |
| 11 | .13 | .04 | .27 | -.02 | .10 |
| 12 | .12 | .24 | .20 | .35 | .19 |
| 13 | .21 | .27 | .25 | .31 | .41 |
| 14 | .16 | .20 | .15 | .49 | .36 |
| 15 | .16 | .19 | .34 | .35 | .41 |
| 16 | .23 | .32 | .29 | .25 | .44 |
| 17 | .17 | .22 | .20 | .30 | .35 |
| 18 | .08 | .25 | .40 | .39 | .08 |

*Note*. The entry for $n$ at the top of each column is the number of usable first-attempt tests that were used to compute the item-total correlations in the column. Shading indicates that the item has an item-total correlation that is negative or less than .10 and therefore needs to be reviewed and possibly revised or replaced.

Table C3

Percentage of Renewal Applicants Who Would Have Passed on the First Test Attempt if Different Passing Cut-Points (Number Wrong) Had Been Used for Each Form of the Spanish DL 5 (Rev. 6/07)

| Number wrong | Form 1 ($n$ = 181) | Form 2 ($n$ = 159) | Form 3 ($n$ = 166) | Form 4 ($n$ = 182) | Form 5 ($n$ = 176) | Total ($N$ = 864) |
|---|---|---|---|---|---|---|
| 0 | 0.6 | 1.3 | 2.4 | 3.8 | 2.8 | 2.2 |
| 1 | 5.0 | 3.8 | 6.6 | 13.2 | 8.0 | 7.4 |
| 2 | 9.9 | 11.9 | 16.3 | 28.6 | 14.2 | 16.3 |
| 3 | 24.3 | 20.8 | 25.3 | 45.6 | 24.4 | 28.4 |
| 4 | 42.5 | 32.7 | 35.5 | 56.6 | 36.4 | 41.1 |
| 5 | 59.7 | 47.8 | 47.0 | 63.7 | 47.7 | 53.5 |
| 6 | 75.1 | 62.9 | 54.2 | 78.0 | 61.9 | 66.8 |
| 7 | 83.4 | 77.4 | 66.3 | 80.2 | 73.3 | 76.3 |
| 8 | 87.8 | 84.9 | 75.9 | 86.3 | 83.0 | 83.7 |
| 9 | 93.4 | 90.6 | 84.9 | 90.7 | 86.4 | 89.2 |
| 10 | 96.1 | 95.6 | 93.4 | 95.1 | 92.0 | 94.4 |
| 11 | 98.3 | 96.2 | 95.2 | 96.7 | 94.9 | 96.3 |
| 12+ | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

*Note*. The entry for $n$ at the top of each column is the number of usable tests that were used to compute the percentages in the column. The shading highlights the pass rates at the current passing standard of three allowable errors.

Table C4

Identification of Problem Items on Each Form of the Spanish DL 5 (Rev. 6/07) Test for Renewal Applicants on the First Attempt

| Problem Indicator | Form 1 Items | Form 2 Items | Form 3 Items | Form 4 Items | Form 5 Items |
|---|---|---|---|---|---|
| Pass rate too high[a] | | 2, 18 | | | |
| Pass rate too low[b] | 12, 13, 18 | 3, 5, 6, 10, 11 | 3, 8, 9, 13, 16 | 17 | 5, 11, 14, 15 |
| Item-total correlation too low or negative[c] | 3, 9, 18 | 11 | | 11 | 18 |
| Distracter selected too often[d] | 13, 18 | 5, 11 | 8, 9, 16 | | 11, 14 |
| Distracter selected too infrequently[e] | 17 | 2, 5, 12, 15, 17, 18 | | 4, 5 | 1 |

*Note*. The entries in the table are the test question numbers printed on the test form. An item may be represented in more than one problem category. The underlining of an item indicates that the item also appeared to be problematic on the English test translation.

[a]More than 95% of applicants answered the item correctly. [b]Less than 60% of applicants answered the item correctly. [c]The item-total correlation was negative or less than .10. [d]The distracter was chosen more often, or almost as often, as the correct answer. [e]A distracter was selected by 2% or fewer applicants.